

Date: March 8, 2012

To: Delaware Justice Reinvestment Task Force

Subject: Data Challenges and Recommendations

From: Vera Institute of Justice, Center on Sentencing and Corrections

Executive Order 27 requires the Delaware Justice Reinvestment Task Force to “undertake a comprehensive review of the state’s criminal justice system and the factors driving jail and prison population growth.”¹ Vera staff provided comprehensive analyses to assist the Task Force in meeting its mandate.

Through our efforts to understand Delaware’s data, Vera staff learned of opportunities to improve Delaware’s capacity and ability to conduct analyses that will support future collaborative criminal justice planning efforts. Crucial information that could assist criminal justice stakeholders in planning and decision making is not readily available. We found that certain data are not captured, that data are missing or cannot be matched between systems, and that different agencies and users handle data inconsistently.

These issues delayed Vera’s analyses and prevented research staff from answering certain questions posed by the Task Force. Despite these challenges, Vera worked closely with the Department of Correction, the Delaware Justice Information System, and the Statistical Analysis to produce careful and thorough analyses upon which the Task Force can rely.

To address these issues, Vera makes the following recommendations:

1. **Convene a data working group.** This group would bring together staff at stakeholder agencies who work directly with the data and create a collaborative forum to discuss data issues and create solutions.
2. **Develop data standards.** One task of the working group should be to create and implement standards for data entry, retention, and validation to ensure reliable and accurate data.
3. **Determine data needs and quality.** The state should conduct a needs assessment and data quality audit of the state’s criminal justice data and systems. This assessment would determine whether current data and systems are providing adequate information to stakeholders for system-wide criminal justice planning and decision making.

The pages that follow detail these issues and recommendations.

Administrative Data Challenges and Recommendations

Two preliminary comments are in order. First, our observations about the data's limitations and our recommendations are based not on working directly with agencies' data systems, but rather on the data Vera received. Second, we relied heavily on guidance from agency staff at the Delaware Justice Information System (DELJIS), Department of Correction (DOC) and Statistical Analysis Center (SAC). Each agency provided valuable advice and assistance as Vera worked to interpret and understand these complex data files. We are extremely grateful for all of the time and effort provided by the staff at each agency.

I. Data Sources

Vera obtained all data on Delaware's pre-sentenced population from DELJIS.² Files included information on type of admission/release, charges, and history of capiases. Vera obtained data from DELJIS and DOC on Delaware's sentenced population.³ DOC data provided basic counts of sentenced admissions, releases, and stock population, as well as sentence length and time served information, while DELJIS provided charge information.

II. Summary of Data Challenges

Vera researchers encountered a number of challenges with the data that impeded our efforts to provide the Delaware Justice Reinvestment Task Force with accurate and detailed information about certain aspects of Delaware's correctional population. Below we describe specific challenges as they relate to data not being adequately captured, missing data, data quality, and inconsistent handling of data.

Data Not Captured

Vera researchers encountered several instances in which Delaware data managers reported that data were not available; three were especially notable. A major challenge was the lack of recidivism figures. Although there is variation across states in definitions of recidivism and how these figures are calculated, basic recidivism rates are crucial in assessing criminal justice system functioning and helping to craft effective policies. We have learned that the SAC will soon be analyzing and reporting recidivism.

While Vera researchers were able to identify those detained or incarcerated for a violation of probation or parole, there was no way in the data to distinguish between technical violations and violations for new offenses.⁴ Stakeholders informed us that standard information systems do not routinely record the information necessary to make such distinctions. Paper violation reports capture this information, but it is not currently fed into or accessible from administrative data systems. We understand that DOC is exploring ways to expand the capacity of the current data system to allow this.

Task Force members expressed strong interest in understanding the impact of failure to appear rates on detention admissions, but data were not available to directly measure this concept. Vera used history of capiases issued as a proxy for failure to appear rates.⁵

Missing Data

Aside from data not being collected to address certain research questions, researchers were also faced with a significant rate of “missing” values for data that were collected. This means that while data fields were designed to capture the information requested, this field was blank for a large share of cases in the files that Vera received.

The lack of data in Vera’s files can be attributed to two factors: (i) cases in DOC data that could not be matched with DELJIS charge data⁶ and (ii) cases in DOC data that were matched with DELJIS charge data, but for which data in DELJIS charge fields were not populated. Our commentary focuses on the latter problem, charge data that were not provided in the DELJIS files we received.

For large proportions of the DELJIS sentenced population data that Vera received, information on offense of conviction, including statute number and description, were missing from the files Vera received (25.8% of stock, 11.1% release, 11.8% of admissions).⁷ Inconsistent entry of data also prevented Vera researchers from accurately assessing the extent of sentenced time accounted for by “time served” sentences.⁸

Data Quality

While working with administrative data, Vera researchers became aware of challenges with the accuracy and consistency of data we received. This section highlights some of these issues. Again, these observations are based only on our experience with the files we received; we did not access the original systems from which the data were pulled.

SENTAC Levels. In analyzing Delaware’s sentenced population data, Vera was interested in establishing counts of admissions to and releases from each of the five SENTAC levels. DOC recommended we rely on the facility “location” information to determine SENTAC level. However, the location record is not a reliable determinant of SENTAC level for a number of reasons. The primary concerns are that one facility can house people at different SENTAC levels and that the system allows records to be open for a single individual at multiple facilities at the same time.⁹

Calculation of Sentence Length. To calculate sentence length, Vera relied on data and guidance provided by the DOC. DOC did not provide overall sentence length information; instead, Vera received data on sentence type and length at the charge level. A judge may order that sentences be served consecutively or concurrently, or suspend a portion of a sentence if certain conditions are met. While the DOC did provide guidance on how sentences are generally applied by the courts (i.e., the Benchbook sentencing policy states that Level 4 and 5 sentences should be consecutive, while Levels 1 through 3 sentences are generally concurrent), DOC data did not include specifics of how charge-

level sentences were to be served. For instance, for an individual admission it was not known whether to sum sentence lengths across offenses, to take the maximum value, or some other alternative. Vera researchers used the guidelines provided by DOC to provide the best estimate of sentence length under these circumstances.

Some of this can be attributed to difficulty in interpreting courts' sentencing orders, while other challenges stem from the limitations of a data system that may not be equipped to accurately capture all of the information in a sentencing order. Our investigation into this issue revealed that multiple agencies must interpret sentencing orders. This is inefficient because it duplicates efforts. From a data quality perspective, it can lead to inconsistent results.

Classification of Charges. Our analysis of DELJIS capias history data revealed that two outdated statutes continue to be used in the data to indicate type of capias. DELJIS staff explained that these statute numbers are used to indicate generic capias (11 Del. Code §5704) and capias for non-support (13 Del. Code §521).

In the DELJIS sentenced data files, charges identified as violations of probation or parole (VOPs) are categorized in the "crime class variable" as misdemeanor offenses. However, such offenses are not misdemeanors by statute, but are a separate class of offenses, and should be categorized in the data to allow for more accurate analysis. To provide an accurate count of felony, misdemeanor and other charges, Vera researchers recoded the crime class variable in DELJIS data to move all violations of probation/parole into a separate category.

Differences between Vera files and CJIS data. There were several examples of instances where the information in the files we received did not match up with information pulled directly from the Criminal Justice Information System (CJIS). For instance, spot checks revealed inconsistencies in the type of admission (while our data showed internal transfers, CJIS data indicated new admissions), inaccurate year of admission (all records that appeared in our files should have been admitted in 2010 but some were shown to actually be admitted in 2011), and incorrect charges being associated with a given case. These inconsistencies may be due, in part, to parameters used to extract data from DACS for Vera, and how records are linked both internally within the DACS system and between DACS and DELJIS.

Counts of Admissions and Releases. The difference between the number of sentenced admissions and releases is crucial in assessing the trajectory of the correctional population. Vera's analysis showed more admissions than releases in 2010, suggesting that the sentenced population is increasing. There may be reasons to doubt the accuracy of this finding.¹⁰ The files Vera received suggest that the population is growing, but DOC's annual reports show a consistent downward trend in the size of the sentenced population the past few years. DOC assured Vera researchers, though, that the files they provided included accurate counts of admissions and releases.

Inconsistent handling of data

Many different agencies—such as the DOC, the courts, the Office of the Attorney General—and numerous users within each agency are responsible for entering and managing data. Understandably, this practice can lead to inconsistencies in the way data are understood, entered, and handled. While some of the issues described above may also be due to these sorts of inconsistencies, here we highlight some additional areas in which inconsistent data handling was especially apparent.

- **Admission/release type.** The codes used to classify types of admissions and releases were used inconsistently. For instance, we learned that the code 460 should be used to indicate a release from detention to another state or a federal authority. We learned, however, that a different code was used for some of these cases. This inconsistency leads to less accurate characterizations of admission and release methods.
- **Out-of-state charges inconsistently identified.** Identifying inmates serving sentences in Delaware for non-state charges presented another challenge. In the data Vera received, there was no consistent way to identify these individuals in either DOC or in DELJIS records owing to practices in data entry. In some cases, the user enters an analogous Delaware statute that is most similar to the out-of-state charge. In other cases, the user may leave the charge field blank or use a generic code indicating a non-Delaware or federal charge. Due to the inconsistent practice, we could not reliably determine whether charges originated from a Delaware state charge or from another jurisdiction. In DELJIS, the variable that Vera initially sought to use to identify out-of-state charges was found, in practice, to be used more broadly to capture any case for which there is no matching charge in the DELJIS system. Often, cases assumed to be as ‘out of state’ according to this variable were actually cases in which the charge information received by DELJIS could not be matched with a charge that was available in the system, but not *necessarily* cases in which the charge originated from out of state. Such cases include civil or family court cases or admissions to DOC pending trial. Based on conversations with stakeholders, we understand that there may be another variable to more reliably identify out-of-state charges, but this variable is rarely used.¹¹
- **Some release dates precede admission dates.** Vera researchers found that some records in the sentenced population have release dates that precede the recorded admission date. Discussions with staff at SAC revealed that this may happen in cases where the sentencing order takes a long time to reach the DOC from the Courts or there is some other data entry delay. We also learned that this frequently happens with time-served sentences. For these cases, DOC back-dates the release date in their data system to the day on which the offender was intended to be released or flowed to a different SENTAC level. In some cases, the admission date reflected in the data system is not the actual date of admission to DOC, but the date on which the sentencing order was received or the data was entered. This impeded Vera’s ability to assess actual time served in incarceration.

- **Matching multiple data sets/Use of identifiers.** Vera became aware of some inconsistencies in data handling as we worked to match DOC and DELJIS files. Matching files using a combination of the individual identifier (SBI number) and the uniform case, or DUC, number revealed challenges with the use of the case number variable in DOC. While DUC numbers in DELJIS follow one of two formats, in DOC data the format varies¹² because they are using this field to store case numbers associated with different systems, including civil court, family court and out-of-state systems. Matching cases on SBI number and DOC ID also revealed inconsistencies in the use of DOC ID in the DELJIS database.

III. Recommendations

1. Convene a Data Working Group

Based on conversations with stakeholders, we understand that a primary reason for data challenges is that DOC's DACS system was developed separately from DELJIS. DELJIS serves as the main repository and source of criminal justice data in Delaware, and DOC must fit its data into static fields in DELJIS. This has led to inconsistent use of certain DELJIS fields, posing some obstacles for research and evaluation.¹³

A first step in addressing this challenge is to create a forum for staff at stakeholder agencies who work directly with the data to facilitate communication among them. In working with DELJIS, DOC, and SAC, we were impressed with the depth of understanding of staff at each agency. While it is important to inform agency administrators about these discussions, this forum should focus on staff who work directly with the data on a regular basis. Due to the complex nature of the data systems, and the fact that staff at each agency are focused primarily on entering data in a manner that is most useful for their purposes, they may not be aware of the challenges other agencies face in working with data they enter. By communicating and collaborating directly, these individuals can readily identify challenges and rely on their experience to suggest responsive solutions that benefit the group as a whole.

2. Develop data standards

One of the working group's initial tasks should be to create and implement standards for data entry, retention, and validation.

Vera encountered numerous instances in which data were found to be inconsistent, many of which are described above. This limited our ability to examine certain aspects of Delaware's criminal justice system. The number and complexity of the systems involved in tracking individuals' movements in the system call for guidance that will ensure more accurate and reliable data.

Those most familiar with the intricacies of working with these systems on a regular basis would be well positioned to recommend processes that could be implemented to enhance

the consistency with which data are entered and implement checks to assesses the quality of data that have been entered. For instance, the challenges described above dealing with inconsistent use of admission and release codes may be an issue that could be resolved with clearer guidelines about which codes should be used under what circumstances and a system of checks to assess whether these guidelines are being followed. Clear policies about how to handle entering admission and release dates could also mitigate the problems related to calculating length of stay. These recommendations would flow naturally from communications between members of the working group to identify challenges in the current system.

3. Determine data needs and quality

Another key step to moving forward would be to conduct a needs assessment and data quality audit of the State's main criminal justice data systems.

The information systems available in Delaware—as in most states' corrections agencies, courts, and other key organizations—were designed for day-to-day operational use. They capture information needed to manage cases on a docket or the population of a prison, to generate required reports, and to meet federal, state, and local requirements. Agencies created their systems to serve their own needs, rarely to gather data for use in cross-system analysis. Policy makers in the past seldom asked analytic questions of the data stored in these systems, so the quality of the data and the ease with which it could be analyzed were not always prioritized.

For these reasons, an audit could review currently captured data, as well as inquire into data and analysis not currently available that would facilitate cross-agency analysis, allowing the State's elected officials, policymakers, and agency leaders to make informed criminal justice policy decisions. Technology has advanced considerably in the last decade, improving usability, automatic quality control mechanisms, reports, and customizable fields. Currently available systems, for case management, for example, provide easily customizable, interactive online access. A data needs assessment may reveal that a new component or system will be helpful to meet the needs of criminal justice stakeholders for system-wide planning and decision making.

Endnotes

¹ Del. Exec. Order No. 27 (July 25, 2011), http://governor.delaware.gov/orders/exec_order_27.shtml.

² We received files on all pre-sentenced admissions and releases in 2010 and the stock population on December 31, 2009. After an additional data request Vera received information on the capias history of 2010 pre-sentenced admissions. Data were received in txt format at the charge level and imported into SPSS version 19 for cleaning, matching and analysis.

³ From DOC, Vera received three flat files in excel format, held at the case-charge level. From DELJIS, Vera received three flat files in txt format. The stock and release files were at the case-charge level, while the admissions file included data on criminal history. Data were imported into PASW version 18 for cleaning, matching, and analysis.

⁴ DOC reported they only maintain these data at the aggregate level, not at the individual-case level. Vera staff reported aggregate data obtained from DOC to the JRI Task Force, but were unable to provide analysis at the case-level. Discussions with SAC revealed that such information is tracked in only more recent entries and thus would not be included for the 2010 population Vera was examining.

⁵ We understand that data systems contain the information that would be necessary to calculate actual rate of failing to appear (as opposed to number of capias issued) but these figures are not normally calculated.

⁶ Using live data systems, SAC was able to locate some of these missing data using various combinations of identifiers, including DOC ID, SBI number and DUC number.

⁷ Communications with staff from DELJIS, DOC and SAC revealed that certain types of cases, namely civil non-support, family court, and out of state cases, will not be found in DELJIS. We learned that some of the missing data could also be accounted for by the fact that when a case is closed with ‘probation before judgment’ (or in a similar manner) there is no information entered in field for “conviction charge”, because there has not been a conviction. Our data did not include any indicators to identify cases where this issue would apply so we were unable to determine the extent to which this method of closing a case accounted for missing conviction charge data.

⁸ Vera was interested in examining how frequently individuals are sentenced to “time served.” However, in the files Vera received 99 percent of cases had a value of 0, leading researchers to question the accuracy of this information. Conversations with SAC revealed that this value is inputted as it comes from the Courts and is dependent upon the sentence effective date. If the sentence effective date is back-dated to account for the total amount of credit for jail time served, the value entered into the variable used to analyze “time served” is 0. We learned that one can infer a sentence composed entirely of “time served” when the effective date of the sentence is the same as the release date, but this practice was not used consistently enough to use this approach as a workaround.

⁹ Individual facilities may house different types of offenders. For example, a level V facility may also house individuals sentenced to Level IV quasi-incarceration, but who are awaiting space in a Level IV facility. Using the “location” variable, this individual would be mis-identified as being held at Level V. Vera found many cases in which a single individual will have multiple admission records for the same admission date open in two different facilities. Location records are used to identify both physical (confinement) and non-physical locations (such as probation or restitution only).

Regarding the issue of multiple locations, SAC shared that this can occur for a number of reasons: (i) DOC allows for records to be opened for individuals flowing from level IV to level III up to three days in advance of the actual transfer of the inmate; (ii) in certain instances, open records are not closed once an inmate transfers locations; and (iii) a large number of probation records remain open for those whose sentences have expired.

¹⁰ Vera received two sets of files on sentenced population releases that provided substantially different counts of the number of releases. DOC recommended that the second file (which included the smaller count of release) was the most accurate. Table 1 summarizes the admission and release figures calculated based on the data Vera received.

Table 1. Summary of Admission and Release Counts

	# of Admissions	# of Releases (Original)	Releases (original) - Admission	# of Releases (Revised)	Releases (revised) - Admissions.
Levels I – III	13,859	14,504	+ 645	14,191	+ 332
Level IV	3,963	3,795	- 168	3,505	- 458
Level V	7,760	7,614	- 146	6,946	- 814

In the initial set of files we received from DOC there were 645 more releases than admissions in Levels I through III, 168 fewer releases than admissions in Level IV and 146 fewer releases than admissions in Level V. After conversations with DOC we learned that we needed additional data that was not in the original files in order to accurately calculate length of stay. DOC obliged our request and re-sent the release files with the new variable. In this second version of the release files there were substantially fewer releases at each level. This decrease in the number of releases also impacted the ratio of the number of releases relative to the number of admissions, and thus impacted our interpretation of the way in which the size of the corrections population may be changing. In the revised release files in Levels I through III there were 332 more releases than admissions, in Level IV 464 fewer releases than admissions, and in Level V 814 fewer releases than admissions. Using the revised release files we see that the number of releases and admissions are more similar in Levels I through III, but the difference in the number of releases and admissions is much greater in levels IV and V. Considering the data provided in the revised release files we would expect that the size of the population in Levels IV and V would increase.

Data from the DOC 2011 annual report, however, shows that the size of the populations in both Levels IV and V have been declining for the past several years. This report reflects a population count on June 30 and uses fiscal years as opposed to calendar years. For these reasons the population counts based on Vera's analysis should not be identical to those in the annual report, but the trends reflected in the DOC annual report are still as useful metric for evaluating the accuracy of the data we received.

¹¹ Vera did receive a variable to indicate whether a case was an ICC charge, but only one admission was coded as an ICC case, leading researchers to question the accuracy of this information.

¹² Examples of these inconsistencies are: CN/CS/CK prefix: according to DOC, these are family court cases for New Castle, Sussex and Kent counties and will not match any case in DELJIS; CP/CR prefix: Interstate or Federal offenses, and will not have a Delaware case; B prefix: family court cases and do not match DELJIS; DACS prefix: indicate civil non-support cases and have no match to DELJIS; Case numbers containing hyphens: preliminary investigations by DOC revealed both federal and civil family court cases, which will not appear in DELJIS; State abbreviations (e.g., AZ, NY) at the end of the case number are out of state cases that have no match in DELJIS; Old case number format, with DI suffix: may not always match case numbers in DELJIS

¹³ For instance, see 'Calculation of Sentence Length' on pg. 3, 'Out of State Charges' on pg 4, and 'Matching multiple data sets/Use of identifiers' on pg. 5.